Adaptive sampling for QoS traffic parameters using fuzzy system and regression model

SALAMA, Abdussalam, SAATCHI, Reza <http://orcid.org/0000-0002-2266-0187> and BURKE, Derek

Published version

SALAMA, Abdussalam, SAATCHI, Reza and BURKE, Derek (2017). Adaptive sampling for QoS traffic parameters using fuzzy system and regression model. International Journal of Mathematical Models and Methods in Applied Sciences, 11, 212-220.

Copyright and re-use policy

Adaptive Sampling for QoS Traffic Parameters Using Fuzzy System and Regression Model

A. Salama and R. Saatchi
Department of Engineering and Mathematics
Sheffield Hallam University
Sheffield, United Kingdom

D. Burke
Sheffield Children's Hospital
Sheffield, United Kingdom

***Abstract*— Quality of service evaluation of wired and wireless networks for multimedia communication requires transmission parameters of packets making up the traffic through the medium to be analysed. Sampling methods play an important role in this process. Sampling provides a representative subset of the traffic thus reducing the time and resources needed for packet analysis. In an adaptive sampling, unlike fixed rate sampling, the sample rate changes over time in accordance with transmission rate or other traffic characteristics and thus could be more optimal than fixed parameter sampling. In this study an adaptive sampling technique that combined regression modelling and a fuzzy inference system was developed. The method adaptively determined the optimum number of packets to be selected by considering the changes in the traffic transmission characteristics. The method's operation was assessed using a computer network simulated in the NS-2 package. The adaptive sampling evaluated against a number of non-adaptive sampling methods gave an improved performance.**

Keywords*— *adaptive sampling; computer network quality of servic; regression modle; fuzzy logic.

I. Introduction

Evaluation of effectiveness of computer networks for communicating various applications is important for allowing network service providers and users to have an improved understanding of how well the services perform against the expectations and ways to identify better allocating resources. This evaluation entails analysing the traffic parameters such as delay, jitter and packet loss ratio that need to be gathered by monitoring information packets [1][2]. However, performing this monitoring in real-time is computationally intensive as a large a number of packets are involved [3].

Sampling is an important process that allows the traffic to be represented by a smaller number of information carrying packets. The process is carefully performed to ensure the transmission attributes of the original traffic are maintained by the selected packets. Sampling can be performed adaptively or in a non-adaptive manner [4] [5]. In an adaptive sampling, the selection process considers the changes in the traffic's behavior such as an alteration in transmission pattern. In nonadaptive sampling the sampling parameters are predefined and do not consider the changes in the dynamics of the traffic [6][7]. Thus adaptive sampling could be more optimal in its performance resulting in better utilisation of resources, reducing processing time and facilitating real-time traffic analysis. Therefore, sampling is an important precursor to quality of service (QoS) assessment for computer networks. The QoS approaches either priorities transmission of specific time-sensitive applications such as video conferencing over other applications such as file transfer or provide a certain level of guarantee to ensure availability of required resources such as bandwidth [4] [5]. The relevance of sampling in computer networks is illustrated in Fig. 1.

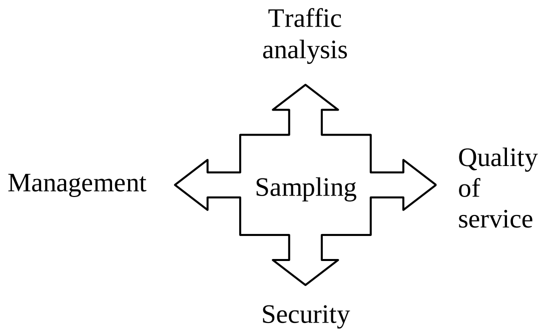


Fig. 1 Role of sampling in computer networks

In this study an adaptive sampling method based on a combination of regression modelling and fuzzy logic was developed and its performance was evaluated for a combined wired and wireless networks. The traffic was initially modeled using regression analysis. Regression analysis is an approach for exploring the relationship between dependent and explanatory variables [8][9]. Regression can be linear or nonlinear but linear regression is commonly used for predictive analysis and is the type used in this study. Regression models has been used for future sensors network readings, allowing network components to be predicted based on current captured data or based on nearest network node [10]. This led to a reduction in the amount of transmitted data packets.

In our study, the output of regression model was interpreted using fuzzy logic. Fuzzy logic uses linguistic rather than numerical values to process information and has an ability to model complex modeling problems more manageably than mathematical formulae. As a result they are becoming increasing useful in network managements involving decision making, control, modelling, security and traffic analysis.

In the conventional (or crisp) logic, a scenario (such as belonging to a group) can either be true (binary 1) or false (binary 0). In fuzzy logic however there is a continuum between true and false as shown in Fig. 2. Therefore, in fuzzy logic there are degrees of membership ranging from 0 to 1 which are defined by membership functions.

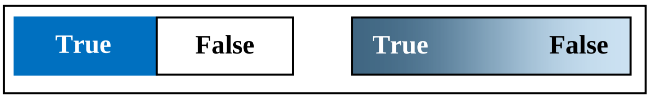


Fig. 2 Conventional and fuzzy logics

A structure to implement fuzzy logic for data analysis is by fuzzy Inference System (FIS) shown in Fig. 3.

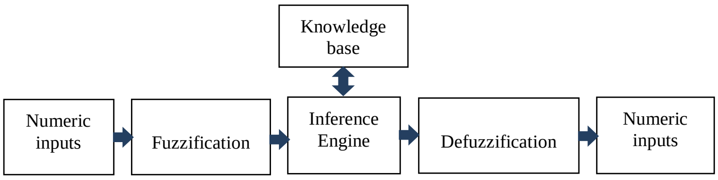


Fig. 3 A block diagram of fuzzy inference system.

The numeric inputs such as the values for packet transmission delay are initially fuzzified through membership functions to determine the degrees they belongs to a set that facilitates a range of values such as *low*, *average* and *high*. The knowledge base represents the domain knowledge (i.e. traffic information) coded by a number of IF-THEN rules. These rules map the fuzzified inputs of the FIS to its fuzzified output. For example a rule may state IF delay is *high* THEN QoS is *poor*. The numeric output for the FIS is obtained through defuzzification process that like fuzzification, uses membership functions [11] [12]. FIS is valuable for computer network traffic sampling as multiple traffic parameters can be suitably combined to suitably interpret changes in traffic behavior [13].

In the following sections, an overview of nonadoptive sampling methods of systematic, random, and stratified is provided. These were used to for comparison of the developed adaptive sampling method. They rely on packet count and tend to have a simple operation [14] [15].

In systematic sampling every n^{th} packets amongst successive groups of k packets are selected. In random sampling, the position of the selected packet is random amongst successive groups of k packets. Stratified sampling is similar to the operation of random sampling. Random numbers are generated and the packets are selected according to their position. New n random numbers are obtained for every run for the same sample size. These approaches are illustrated in Fig.4.

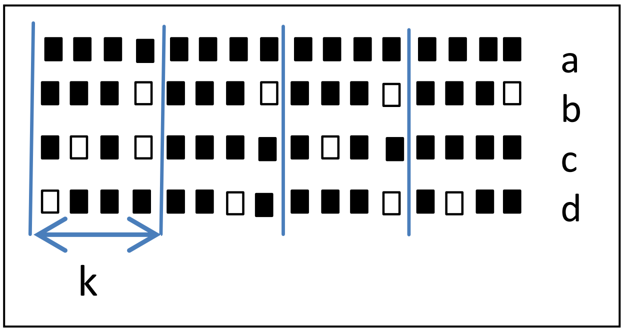


Fig.4. Illustrations of non adaptive sampling techniques, white squares are selected packet during the sampling (a) original traffic, (b) systematic sampling, (c) random sampling and (d) stratified sampling.

II. Related Work

A study explored packet sampling selection schemes, selection trigger and identifying granularity in sampling and proposed a general-purpose architecture to sustain the development of flexible sampling systems [15]. However, working algorithms are still being developed. An OpenFlow solution which provided statistics collection mechanism of a flow level from the data plane was proposed [3]. The proposed PayLess mechanism was a monitoring framework for Software Defined Networking to simplify network management by separating the central controller (control plane) from the data switches (data plane). The solution defined monitoring accuracy, timeliness and network overhead. The proposed PayLess delivered a flexible statistical data flow gathering at different aggregation levels. Their solution used an adaptive statistical collection algorithm which provided accurate information in real-time without adding significant network overhead. The proposed mechanism was demonstrated in Mininet to evaluate its effectiveness. Adaptive sampling techniques based on traffic's statistics were proposed [12] [16]. The techniques adaptively adjusted sampling interval between consecutive sections according to the changes in the measured statistics. However, these techniques used a single traffic parameter such as pack transmission delay.

In our earlier study, an adaptive sampling technique was developed that utilised linear regression for traffic modeling and a fuzzy inference system for data interpretation [6]. It dynamically adjusted the inter-sampling interval (*isi*) by two consecutive sampled sections. However, the technique was using one traffic parameter at a time. This study is a significant further development of adaptive sampling that simultaneously considers three main traffic parameters, namely delay, jitter and packet loss ratio.

III. Methods

A modular and scalable network was designed using a network simulation package called NS-2. The network's design (shown in Fig.5) was based on the recommended hierarchical network structure that divides the network in into three tiers called core, distribution and access. This design improves network management by ensuring its modularity [17] and was compliant with the Open Source Interconnection (OSI) network model [18].

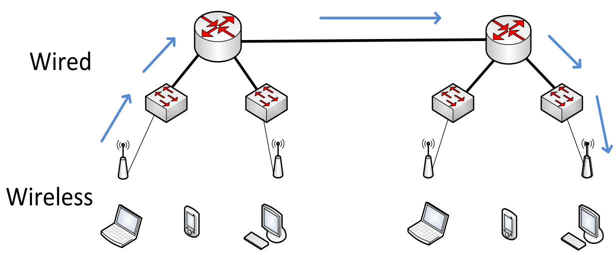


Fig.5. The network design

The wired part of the network contained the core layer and had a capacity of 10 Mbps. The wireless parts contained the distribution and access tiers and were configured in the IEEE 802.11e protocol with Enhanced Distributed Channel Access (EDCA). The wireless channel capacity was 2 Mbps. The routing protocol in this scenario was Destination-Sequenced Distance Vector (DSDV) and the queuing mechanism for all scenarios was First-In-First-Out (FIFO). The queue size was 50 packets.

The transmitted traffic were video streaming, VoIP, HTTP and FTP. The packet size for VoIP was 160 bytes. G711 protocol was used as audio coding with 64 kbps transmission rate. The packet size for video streaming was 512 bytes. The video streaming frames were configured with maximum length of 1024 bytes and MPEG-4 coding scheme. The NS-2 sampling evaluation scenarios ran for 800 seconds. Following each simulation, a trace file was produced by NS-2 that contained the network and traffic transmission details such as the packet types (i.e. data, routing, etc.), transmitted and received times and packet sizes and delivery status. A Perl language based tool was developed to read the information from the trace file and determined the traffic parameters: delay, jitter, and packet loss ratio. These measurements were performed using equations explained below.

Delay (D_i) for the i^{th} packet was determined as in equation (1) where R_i and S_i are the times a packet was received and sent respectively.

$$D_i = R_i - S_i \tag{1}$$

Jitter (J_i) was determined using equation (2) where D_i and D_{i-1} are the delays associated with the current and previous packets respectively. The absolute parameter ensures jitter values remain positive.

$$J_i = absolute\ (D_i - D_{i-1}) \tag{2}$$

The percentage packet loss ratio ($\%PL_i$) was determined by using equation (3) where R_i and S_i are i^{th} packets received and sent respectively.

$$\%PL_i = \left(1 - \frac{\sum R_i}{\sum S_i}\right) \times 100 \tag{3}$$

II. Description of Adaptive Sampling Method

The algorithm used regression to model traffic by considering delay, jitter and percentage packet loss ratio. The output of the model was then interpreted by the fuzzy inference system to adapt traffic sampling. The operation of the algorithm is shown in Fig.6 and its key parameters and elements are explained below. Fig.7 complements the flowchart in illustrating the sampling operation.

- *Pre and post-sampling sections*: These intervals contain the traffic that needs to be sampled. These intervals are kept fixed (predefined) and do not changed during sampling.
- *Inter-section Interval of data packets (isi)*: This interval is between pre- and post-sampling sections. Its duration is adaptively determined by considering the output of the fuzzy inference system.
- *Traffic matrix:* The traffic parameters were represented by an $n \times n$ traffic matrix to form the regression model, where n is the number of subsections in the pre- and post-sampling sections. Each subsection contained n packets. The data modelling was performed for the measured traffic parameters, i.e. delay, jitter and percentage packet loss ratio.
- *Traffic difference calculation using Euclidean distance (ED)*: ED measure was used to determine the amount of traffic difference (*td*) between pre- and post-sampling sections for all traffic parameters.
- *Fuzzy inference system*: FIS was used to determine updated (*isi)* based on the current *(isi)* and the three traffic difference *(td*) values of delay, jitter and percentage packet loss ratio.

The algorithm updates the *isi* length and the pre- and post-sampling sections are determined at the end of each iteration.

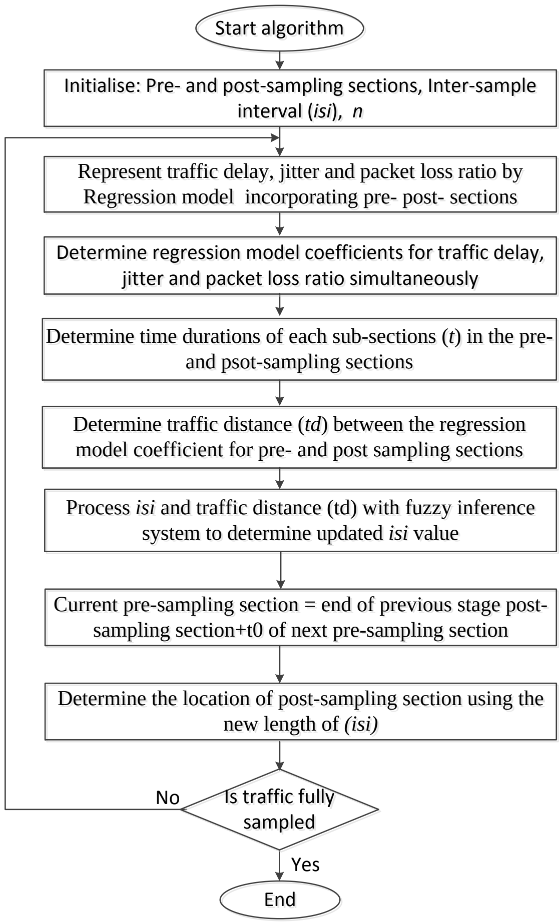


Fig. 6 Flow chart for the adaptive sampling method

The traffic parameters delay, jitter and packet loss ratio were considered as the independent variables representing *p* values in regression equation (4). The sampling section was divided to subsections (s_1, s_2,...s_n). Each subsection contained *(n-1)* packets as shown in Fig.5, where the traffic parameter values for each subsection were entered as a row of the traffic matrix *P* and the associated time period of every subsection were represented by the vector T. The vector $[E]' = [e_1, e_2, \ldots, e_n]'$ is the error vector (the symbol ' signifies transpose).

$$T = PC + e = \begin{bmatrix} 1 & P_{11} & P_{12} & ..P_{1(n-1)} \\ 1 & P_{21} & P_{22} & ..P_{2(n-1)} \\ : & : & : & : \quad : \\ 1 & P_{n1} & P_{n2} & ..P_{n(n-1)} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ : \\ c_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ : \\ e_n \end{bmatrix} \qquad (4)$$

In this study, n was 4 which resulted in 4 sub-sections: S_{1pre}, S_{2pre}, S_{3pre} and S_{4pre} for pre-sampling section and S_{1post}, S_{2post}, S_{3post} and S_{4post} for post-sampling sections as illustrated in Fig.6. Each subsection contained 3 data packets. For both post and pre-sampling sections a 4×4 traffic matrix was formed where each of its rows contained the traffic information of each sub-sections. This was repeated for the pre and post-sampling sections.

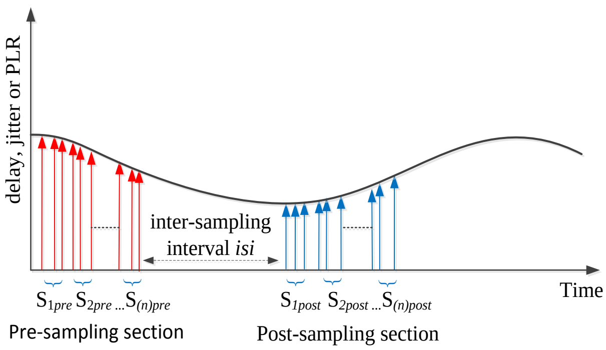


Fig. 7 Traffic representation for the algorithm

The time durations of the subsections were represented by t_1, t_2 ... t_n. These durations were measured by subtracting the arrival time of the last packet for a section from the arrival time of the first packet for the same section. The error vector (e) in tested scenarios was set to zero. The regression coefficients; c_0, c_1 ... c_{n-1} were determined by equation 5.

$$C = P^{-1}T \qquad (5)$$

The magnitude of traffic difference (td) between the pre- and post- sampling sections was determined by comparing their respective regression model coefficients using the Euclidean distance measure as shown in equation 6.

$$traffic\ difference\,(td) = \sqrt{\left(c_{1pre} - c_{1post}\right)^2 + \left(c_{2pre} - c_{2post}\right)^2 + \ldots + \left(c_{npre} - c_{npost}\right)^2} \qquad (6)$$

The fuzzy inference system received the current value of inter-sampling interval (isi) and the traffic difference (td) for traffic parameters delay, jitter and percentage of packet loss ratio and determined the updated value for inter-sampling interval (isi) as shown in Fig.8.

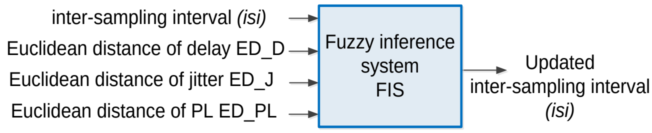


Fig. 8 FIS system to update inter-sample interval

The Mamdani type of Fuzzy Inference System (FIS) was used to dynamically adjust the length of isi. Four inputs were fed into the FIS. They were the current inter-sampling interval, network parameters delay, jitter and packet loss ratio. The inputs and the output were fuzzified using the Gaussian membership functions that has a concise notation and is smooth. The Gaussian membership function is represented by formula is expressed in (7) where c_i and σ_i are the mean and standard deviation of the i^{th} fuzzy set A^i [2].

$$\mu_A{}^i\,(x) = \exp\left(-\frac{(c_i - x)^2}{2\sigma_i^2}\right) \qquad (7)$$

The inputs to the fuzzy inference system, the values of traffic difference for delay, jitter and percentage packet loss ratio and the inter-sampling interval (isi) were individually fuzzified by five membership functions. The traffic difference for delay, jitter and packet loss were represented by *VLow*, *Medium*, *High* and *Vhigh* fuzzy sets. The input inter-sampling interval (isi) was represented by *Vsmall, small, Medium, Large* and *Vlarge* fuzzy sets. The output was defizzzified by four membership functions, represented by *IL* (Low Increase), *NC* (no change), *DL* (Low Decrease), and *DH* (High decrease). These membership functions are shown in Fig.9.

Tables (I) and (II) show the values of membership function parameters for fuzzy inputs (i.e. delay, jitter, PLR, and current isi) and fuzzy output (i.e. updated isi) respectively.

Table I Input membership function (mean delay, jitter and %PLR) and their values

Membership names	**Values**
Very low	0
Low	1.25
Medium	2.5
High	3.75
Very high	5

Table II Mean inter-sample interval difference and output updated inter-sample interval fuzzy membership functions.

Membership functions	**Membership functions**	**Current and updated *isi***
Very small	Decrease low (DL)	0
Small	Decrease High (DH)	25
Medium	No change (NC)	50
Large	Increase low (IL)	75
Very large	Increase high (IH)	100

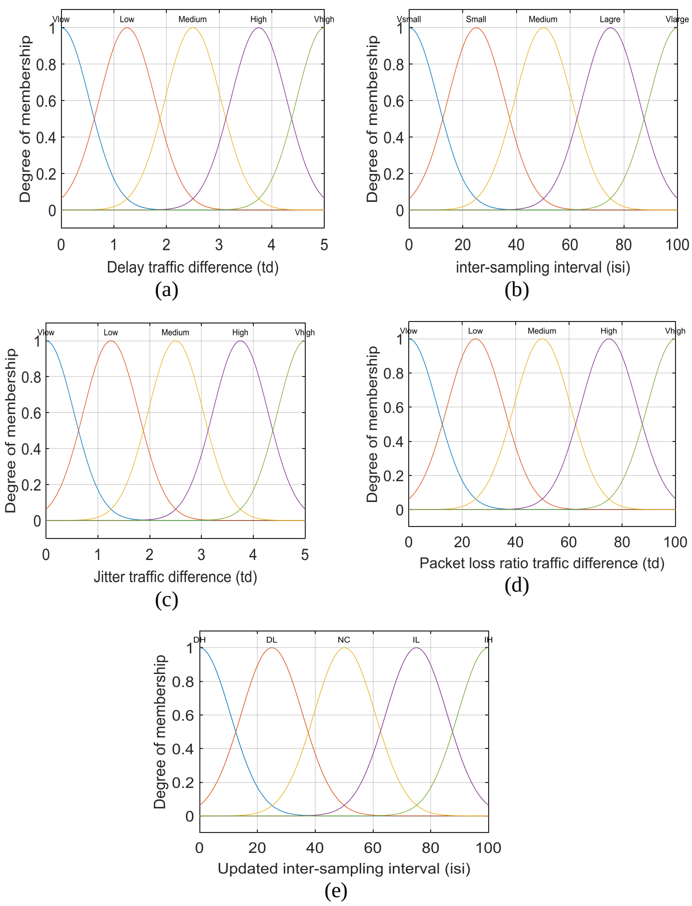


Fig. 9 Membership functions for (a-c) traffic difference sets for delay, jitter and percentage packet loss ratio. (d) inter-sampling interval (e) the updated inter-sampling interval

The implication and aggregation (Fuzzy reasoning) methods used minimum-maximum operation. In this method each rule is applied to the related membership function and the minimum is mapped into corresponding output membership function. The output of fuzzy set from the implication process for each fuzzy rule is combined together by aggregation process to produce one fuzzy set. In this study, the fuzzy output was produced from aggregated fuzzy set (defuzzification) by using the centroid scheme. The centriod scheme uses (9) to return the centre of area under the curve of the aggregated output values [1].

$$Y = \sum_{i=1}^{m} y_i \times \mu_i / \sum_{i=1}^{m} \mu_i \tag{9}$$

where y_i is the centriod of fuzzy region i, m is the number of fuzzy sets obtained after implication , and μ_i is membership value.

In order to assess the efficiency of the developed sampling technique, a comparison of the original data populations to its sampled version was performed. Measurements and comparisons of mean and standard deviation of the sampled packets may not be enough to evaluate the accuracy of sampled version in terms of representing the original data population as they can be affected by outliers [12] [14]. Therefore additional criteria were used to assess the efficiency of the developed sampling technique. The bias indicates how far the mean of the sampled data lies from the mean of the original data [19]. Bias is the average of difference of all samples of the same size. The bias was calculated as

$$Bias = \frac{1}{N}\sum_{i=1}^{N} M_i - M \tag{10}$$

Where N is the number of simulation runs, M_i and M are the means of the traffic parameters for the original population and its sampled version.

Relative Standard Error (RSE) is another criteria used to assess the efficiency of the method, RSE examines the reliability of sampling. RSE is defined as a percentage and can be defined as the standard error of the sample (SE) divided by the sample size (n) as

$$RSE = \frac{SE}{n} \times 100 \tag{11}$$

where n is sample size, SE *is* standard error values of the original and sampled traffic parameters, i.e. (delay, jitter and percentage packet loss ratio) and sampled packets.

Curve fitting is another criterion used to illustrate the behavior of sampled data in terms of representing the original data population. It examines the trend of sampled data versus its equivalent original data by applying the curve fitting. Curve fitting is a suitable tool for demonstrating a data set in a linear, quadratic or polynomial fashion [20] [21]. Curve fitting of data is based on two functions, polynomial evaluation function and polynomial curve fitting function, which can quickly and easily fit a polynomial to a set of data points. The general formula for a polynomial is shown as

$$f(x) = a_0 x^N + a_1 x^{N-1} + a_2 x^{N-2} + \ldots\ldots + a_{N-1} x + a_N \tag{12}$$

The degree of a polynomial is equal to the maximum value of the exponents (N), ($a_0 \ldots a_N$) is a set of polynomial coefficients and (x) is a set of data. Polynomial curve fitting function measures a least squares polynomial for a given data set of (x) and generates the coefficients of the polynomial which can be used to illustrate a curve to fit the data according to the specified degree (N). The polynomial evaluation function examines a polynomial for a given set of data (x) values and then produces a curve to fit the data based on the coefficients that were found using the curve fitting function [20] [22].

Sampling fraction is the proportion of a population that will be counted. Sampling fraction is the ratio of the sampled size (n) divided by the population size (N).

Fuzzy rules processed the values of inter-sampling interval (*isi*), traffic differences for delay, jitter and percentage packet loss ratio to update the inter-sampling interval (*isi*). Table III shows examples of the fuzzy rules.

Table III Examples of fuzzy rules

no	current isi	TD delay	TD jitter	TD packet loss ratio	updated isi
1	Very small	Very low	Very low	None	Increase high (IH)
2	Very small	Very low	None	Very low	Increase high (IH)
3	Very small	None	Very low	Very low	Increase high (IH)
4	None	Very low	Very low	Very low	Increase high (IH)
5	None	Low	Low	Low	Increase low (IL)
6	Small	None	Low	Low	Increase low (IL)
7	Small	Low	None	Low	Increase low (IL)
8	Small	Low	Low	None	Increase low (IL)
9	Medium	Medium	Medium	None	No change (NC)
10	Medium	Medium	None	Medium	No change (NC)
11	Medium	None	Medium	Medium	No change (NC)
12	None	Medium	Medium	Medium	No change (NC)
13	None	High	High	High	Decrease low (DL)
14	Large	None	High	High	Decrease low (DL)
15	Large	High	None	High	Decrease low (DL)
16	Large	High	High	None	Decrease low (DL)
17	None	Very high	Very high	Very high	Decrease low (DH)
18	Very large	None	Very high	Very high	Decrease low (DH)
19	Very large	Very high	None	Very high	Decrease High (DH)
20	Very large	Very high	Very high	None	Decrease High (DH)

td: traffic difference, measured by Euclidean distance

Traffic parameters delay, jitter and packet loss ratio were measured by equations 1-3. The simulation ran for 800 seconds. The linear regression model shown in equation (4) was used to model traffic parameters for pre- and post-sections of the inter-sampling interval (*isi*). The traffic parameter differences were measured using equation (6) to determine the magnitude of the traffic changes. This was performed for the three traffic parameters simultaneously. Fuzzy inference system updated the inter-sampling interval based on the current value of inter-sampling interval, the extent of traffic change and fuzzy rules for each update. The results are shown in Figs.9-10.

IV. RESULTS

As an example, Fig.10 (a) indicates the adaptive updating of *isi* based on the variations in packet delay. Fig. 10(b) indicates the manner the Euclidean distance, the variation of Euclidean distances of delay, jitter and packet loss ratio affect *isi* changes. When variations are large *isi* decreases and vice versa. Fig.10(c) shows the original delay and its trend and Fig.10(d) the sampled delay and its trend. The trends for the original delay and its sampled version are close.

In Fig.11(a)-(d) indicates the manner the developed adaptive sampling method tracked the jitter and percentage packet loss ratio (PLR). In Figs 11(a)-(b) the Euclidean distance are shown. The Euclidean distance of the packet loss ratio variation changed more than the variation of delay and jitter due to rapid change of packet loss ratio, these variations in the Euclidean distance caused the changes in the of *isi* values. In Fig11(c)-(f) the actual (original) jitter and PLR are shown their sampled version. For both traffic parameters, the trends for the original traffic parameters are close to the sampled version.

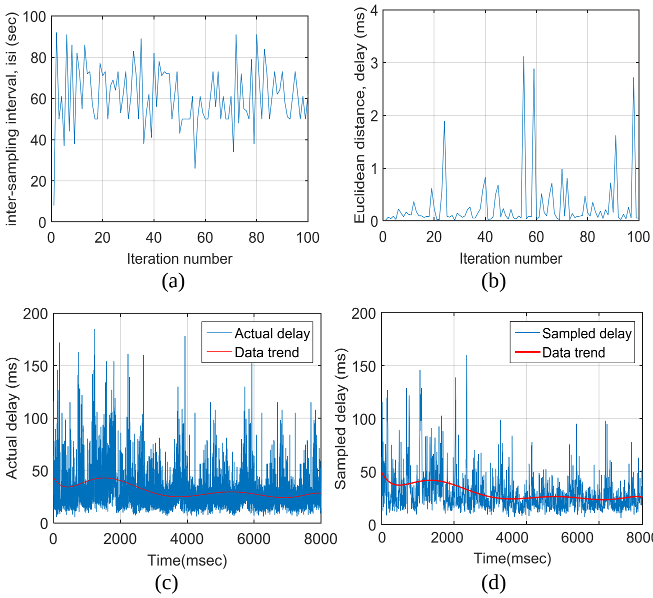


Fig. 10 Typical results obtained from the developed adaptive technique (a) FIS output for the inter-sampling interval (*isi*) (b) traffic difference for delay (c) original traffic delay (d) sampled traffic delay.

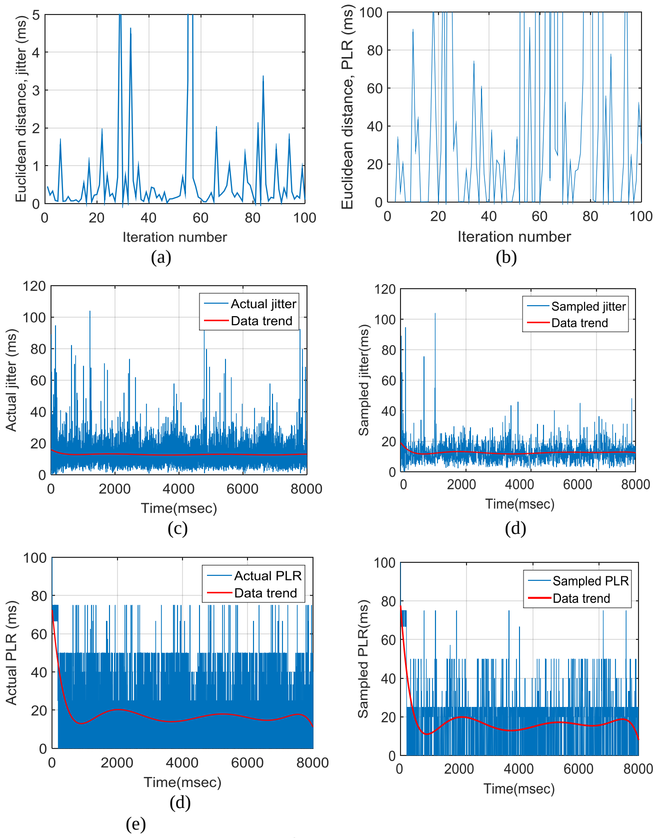


Fig. 11 typical results obtained from the developed adaptive technique: (a) measured traffic difference for jitter, (b) measured traffic difference for packet loss (c) original traffic jitter, (d) sampled traffic percentage jitter (e) original traffic packet loss ratio (f) sampled traffic packet loss ratio.

Table IV provides a summary of delay sampling results for the original traffic (0% sample fraction) and a number of different sample fractions for the adaptive and nonadaptive sampling methods of systematic, random and stratified. Similar information is provided for jitter and PLR in Tables V and VI. To compare the developed adaptive sampling and nonadaptive sampling methods, the bias and relative standard errors (RSE) were determined. They indicate that the developed adaptive method has the lowest relative error and bias values in most of sample fractions as compared as compared with the non-adaptive methods, signifying an improved performance.

Table IV Measurement results of delay using different sampling methods: adaptive, systematic, random and stratified

Unit	Sample fractions %				
	0.0	**9.9**	**13.05**	**20.25**	**29.47**
Adaptive sampling method					
Mean	31.00	30.81	30.80	31.34	30.98
Std.	20.50	22.11	18.60	21.22	20.10
Bias	0	0.188	0.200	-0.338	0.018
RSE	0	0.001	5.97E-04	3.25E-04	4.41E-05
Systematic sampling					
Mean	31.00	30.47	30.77	30.64	30.93
Std.	20.50	20.33	20.29	19.97	20.68
Bias	0	0.53	0.22	0.357	0.072
RSE	0	8.99E-04	6.42E-04	3.12E-04	2.31E-04
Random sampling					
Mean	31.00	31.80	31.37	30.41	31.06
Std.	20.50	21.75	21.39	19.67	20.19
Bias	0	-0.795	-0.373	0.590	-0.055
RSE	0	9.76E-04	6.34E-04	3.02E-04	1.76E-04
Stratified sampling					
Mean	31.00	31.33	30.49	31.08	31.37
Std.	20.50	19.32	21.38	21.22	21.06
Bias	0	-0.333	0.509	-0.081	-0.367
RSE	0	8.87E-04	6.29E-04	3.29E-04	1.86E-04

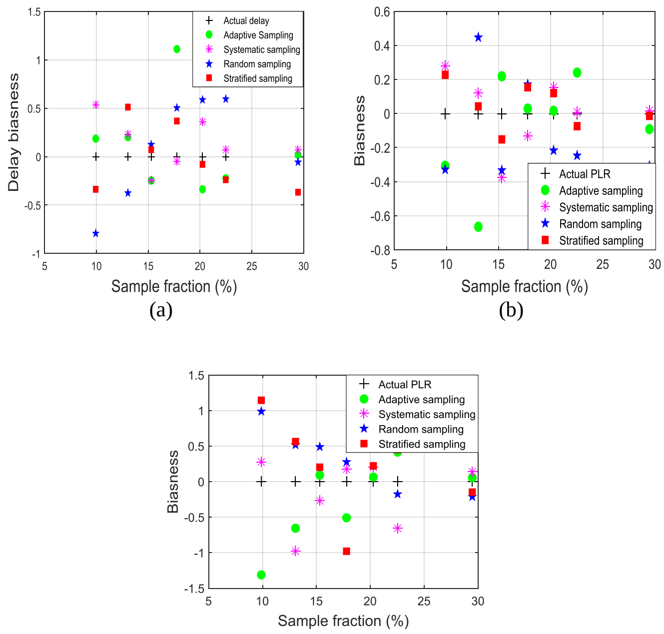
Table 0 Measurement results of jitter using different sampling methods: adaptive, systematic, random and stratified

Unit	Sample fractions %				
	0.0	**9.9**	**13.05**	**20.25**	**29.47**
Adaptive sampling method					
Mean	12.83	12.79	13.50	12.82	12.85
Std.	7.31	7.78	8.44	7.53	7.12
Bias	0	0.040	-0.662	0.0150	-0.015
RSE	0	4.46E-04	2.71E-04	1.10E-04	6.42E-05
Systematic sampling					
Mean	12.83	12.56	12.72	12.68	12.82
Std.	7.31	7.67	6.76	7.37	6.92
Bias	0	0.279	0.118	0.154	0.0164
RSE	0	3.39E-04	2.14E-04	1.15E-04	5.03E-05
Random sampling					
Mean	12.83	13.17	12.39	13.05	13.14
Std.	7.31	8.18	6.18	8.08	7.83
Bias	0	-0.330	0.447	-0.217	-0.305
RSE	0	3.67E-04	1.83E-04	1.24E-04	6.84E-05
Stratified sampling					
Mean	12.83	13.14	12.79	12.71	12.93
Std.	7.31	8.62	7.78	7.21	7.34
Bias	0	-0.306	0.040	0.120	-0.0912
RSE	0	0.0177	0.00317	0.00942	0.00117

Table 0I Measurement results of packet loss ratio using different sampling methods: adaptive, systematic, random and stratified

Unit	Sample fractions %				
	0.0	**9.9**	**13.05**	**20.25**	**29.47**
Adaptive sampling method					
Mean	17.87	17.59	18.53	17.81	17.82
Std. deviation	17.83	18.16	18.23	18.26	17.75
Bias	0	0.277	-0.656	0.064	0.056
RSE	0	8.03E-04	5.85E-04	2.67E-04	3.90E-05
Systematic sampling					
Mean	17.87	19.19	18.85	17.67	17.73
Std.	17.83	19.71	17.64	18.12	0.1426
Bias	0	-1.315	-0.974	0.199	0.142
RSE	0	0.0010204	5.58E-04	2.83E-04	1.29E-04
Random sampling					
Mean	17.87	16.88	17.36	17.29	18.08
Std.	17.83	16.97	17.65	17.53	18.14
Bias	0	0.992	0.513	0.581	-0.208
RSE	0	7.62E-04	5.23E-04	2.69E-04	1.58E-04
Stratified sampling					
Mean	17.87	16.73	17.31	17.65	18.03
Std.	17.83	17.90	17.16	17.89	18.22
Bias	0	1.145	0.566	0.221	-0.149
RSE	0	8.22E-04	5.05E-04	2.77E-04	1.61E-04

Fig. 12(a)-(c) show respectively the comparison of the bias of sampled delay, jitter and PLR from the actual delay, jitter and PLR for different sample fractions using the proposed adaptive sampling method and non-adaptive systematic, random and stratified. The results indicate that the bias was decreased and became closer to zero for all sampling methods when the sample size increased. The results indicate that the proposed adaptive sampling method has a lower bias as compared with systematic, stratified, and random sampling approaches. For example, at 29.47% sample fraction, the bias of sampled delay was 0.018, while the bias values by systematic, random and stratified sampling were 0.072, -0.055, and -0.367 respectively.



(a) (b)
(c)

Fig. 12 Comparisons of biasness of (a) delay, (b) jitter and (c) PLR between developed technique and non-adaptive methods

In Figs. 13(a)-(c) the RSE for sampled delay, jitter and PLR for nonadaptive sampling approaches (systematic, random and stratified) are compared with the measured RSE for the proposed adaptive sampling method. The results indicate the proposed adaptive sampling method has a lower RSE as compared with the nonadaptive sampling approaches. For example, at 29.47% sample fraction, the RSE of sampled delay was 4.41E-05, while the bias values by systematic, stratified, and random sampling were 2.31E-04, 1.76E-04, and 1.86E-04 respectively.

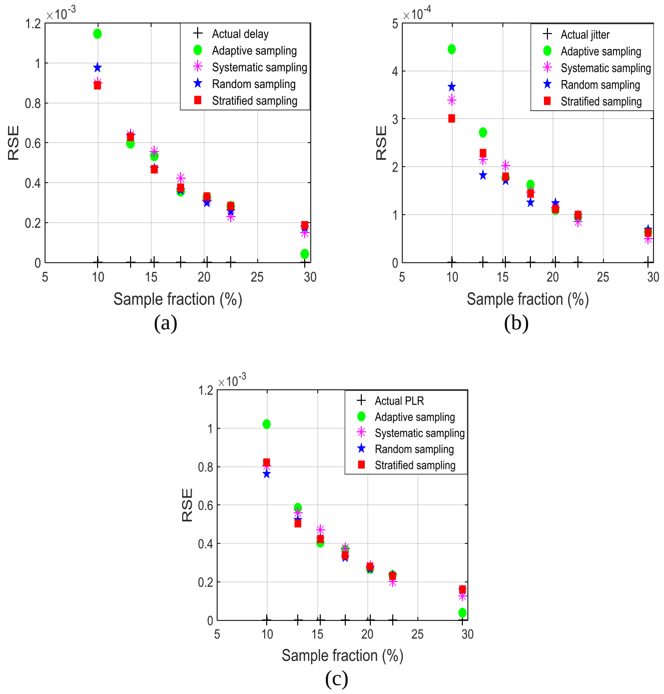


Fig. 13 Comparisons of RSE of (a) delay, (b) jitter and (c) PLR between developed technique and non-adaptive methods

V. Conclusions

A novel adaptive method that sampled multimedia network traffic has been developed and evaluated. It incorporates the traffic parameters delay, jitter and percentage packet loss ratio simultaneously in its analysis. Its performance was compared with the nonadaptive sampling techniques of systematic, random, and stratified. The developed method adaptively increased the inter-sampling interval section resulting in an increase in the number of packets sampled when the traffic variations increased and vice versus. The adaptive sampling method represented the original traffic more accurately than the non-adaptive methods.

VI. Acknowledgment

The authors are grateful to receive Sheffield Hallam University Vice Chancellor's PhD Studentship funding that allowed this work to be carried out.

References

[1] Y. A. Al-Sbou, R. Saatchi, S. Al-Khayatt, R. Strachan, M. Ayyash, and M. Saraireh, "A novel quality of service assessment of multimedia traffic over wireless ad hoc networks", In Next Generation Mobile Applications, Services and Technologies, NGMAST'08, IEEE, The Second International Conference on, 479-484, 2008.

[2] M. Saraireh, R. Saatchi, S. Al-Khayatt, R. Strachan, "Assessment and improvement of quality of service in wireless networks using fuzzy and hybrid genetic-fuzzy approaches", Artificial intelligence review, 27(2-3), 95-111, 2007.

[3] S. R. Chowdhury, M. F Bari, R. Ahmed and R. Boutaba, "Payless: A low cost network monitoring framework for software defined networks", In Network Operations and Management Symposium (NOMS), 2014 IEEE (pp. 1-9). IEEE, 2014.

[4] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, "An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors", IEEE Transactions on Instrumentation and Measurement, 59(2), 335-344, 2010.

[5] J. M. C. Silva, C. Paulo and R. L. Solange, "Inside packet sampling techniques: exploring modularity to enhance network measurements." International Journal of Communication Systems 30.6, 2017.

[6] A. Salama, R. Saatchi and D. Burke, "Adaptive sampling technique for computer network traffic parameters using a combination of fuzzy system and regression model", In: 2017 4th International Conference on Mathematics and Computers in Science in Industry, MCSI 2017, Island, Greece, August. IEEE. (In Press), 24-26, 2017.

[7] A. Salama, R. Saatchi and D. Burke, "Adaptive sampling technique using regression modelling and fuzzy inference system for network traffic", In: CUDD, Peter and DE WITTE, Luc, (eds.) Harnessing the power of technology to improve lives. Studies in Health Technology and Informatics (242). IOS Press, 592-599, 2017.

[8] J. Fan, Y. Liao, and H. Liu, "An overview of the estimation of large covariance and precision matrices", The Econometrics Journal, 19(1), C1-C32 [4], 2016.

[9] J. J. Faraway, Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Vol. 124. CRC press, 2016.

[10] B. Zhang, Y. Liu, J. He and Z. Zou, "An energy efficient sampling method through joint linear regression and compressive sensing", In Intelligent Control and Information Processing (ICICIP), 2013 Fourth International Conference on, IEEE, 447-450, 2013.

[11] A. Egaji Jafari, S. Al-Khayatt and A. Dogman, "Performance evaluation of IEEE 802.11 p for vehicular communication networks", Communication Systems, Networks & Digital Signal Processing (CSNDSP), 2012 8th International Symposium on, 1-5, 2012.

[12] A. Dogma, R. Saatchi, S. Al-Khayatt, and H. Nwaizu, "Adaptive statistical sampling of voip traffic in wlan and wired networks using fuzzy inference system", 2011 7th International Wireless Communications and Mobile Computing Conference, 1731-1736, 2011.

[13] O. A. Egaji, A. Griffiths, , M. S. Hasan, and H. Yu, "A comparison of mamdani and sugeno fuzzy based packet scheduler for MANET with a realistic wireless propagation model", International Journal of Automation and Computing, 12(1), 1-13, 2015.

[14] T. Zseby, "Comparison of sampling methods for non-intrusive SLA validation", In Proceedings of the Second Workshop on End-to-End Monitoring Techniques and Services (E2EMon), 2004.

[15] J. M. C. Silva, C. Paulo and S. R. Lima. "Inside packet sampling techniques: exploring modularity to enhance network measurements." International Journal of Communication Systems 30.6 (2017).

[16] A. Dogma, R. Saatchi, S. Al-Khayatt, "An adaptive statistical sampling technique for computer network traffic. Communication

Systems Networks and Digital Signal Processing (CSNDSP)", 2010 7th International Symposium on, 479-483, 2010.

[17] R. de Souza Couto, S. Secci, M. E. M. Campista, and L. H. M. K. Costa, "Reliability and survivability analysis of data center network topologies", Journal of Network and Systems Management, 24(2), 346-39, 2016.

[18] J. R. Díaz Santos, "Design and Implementation of a Communication Protocol to Improve Multimedia QoS and QoE in Wireless Ad Hoc Network", PhD thesis; Universitat Politècnica de València, http://www.tdx.cat/handle/10803/391793, last accessed 29/09/2017.

[19] T. Zseby, "Comparison of sampling methods for non-intrusive SLA validation", In Proceedings of the Second Workshop on End-to-End Monitoring Techniques and Services, (E2EMon), 2004.

[20] X. Wan, et al. "A T-wave alternans assessment method based on least squares curve fitting technique", Measurement 86 (2016): 93-100.

[21] Guruswami, Venkatesan, and David Zuckerman, "Robust Fourier and polynomial curve fitting." Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on. IEEE, 2016.

[22] G. Lindfield and J. Penny, "Numerical methods: using MATLAB. Academic Press", 2012.